

## MORE TOPICS IN PROBABILITY: INCLUSION-EXCLUSION AND CONDITIONAL PROBABILITY

These notes follow on from the ‘Fundamentals of Probability’, so please read those first.

---

Last time we saw that probabilistic events  $A$  and  $B$  belong to a sample space  $\Omega$  with probability functions  $P(A) \in [0, 1]$  and  $P(B) \in [0, 1]$ <sup>1</sup>. If  $A$  and  $B$  are mutually exclusive, then

$$P(A \cup B) = P(A) + P(B).$$

Furthermore, if  $A$  and  $B$  are independent, then

$$P(A \cap B) = P(A)P(B).$$

These notes will realise such expressions without any assumptions on the events  $A$  and  $B$ . To do so, we use Venn diagrams, very useful tools when talking about probability - there’s a good reason for their popularity!

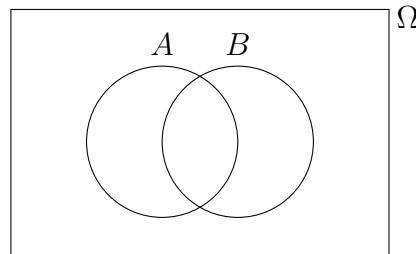


Figure 1: A Venn diagram. The bounding rectangle represents the sample space  $\Omega$ ; all events must be contained within it. The events  $A$  and  $B$  are depicted as circles, with the ‘intersection’ (overlap) representing  $A \cap B$ .

An example of a Venn diagram is shown in Figure 1. Using a Venn diagram is as simple as shading the region you’re interested in; see Figure 2.

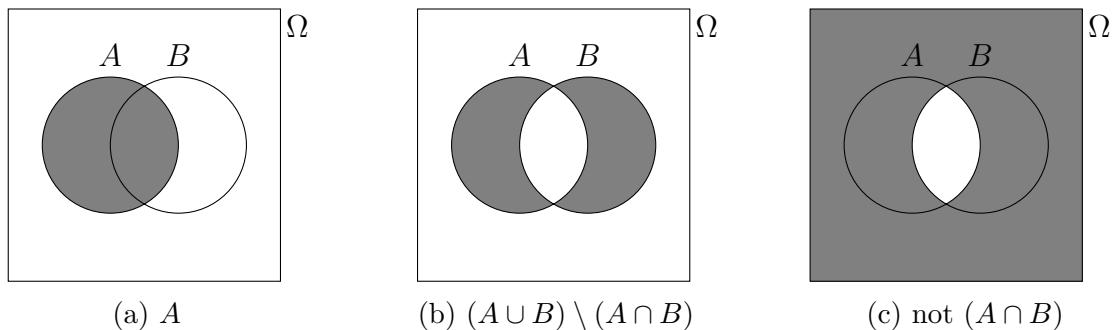


Figure 2: Three Venn diagrams for different combinations of the events  $A$  and  $B$ .

---

<sup>1</sup>We call  $[a, b]$  the *closed* interval from  $a$  to  $b$ . Saying that some number  $x$  belongs to this interval, i.e.  $x \in [a, b]$ , is equivalent to the inequalities  $a \leq x \leq b$ . Therefore  $P(A) \in [0, 1]$  means  $0 \leq P(A) \leq 1$ .

The operator  $\setminus$  seen in Figure 2(b) is the ‘set minus’ operator. It is the set version of subtraction, meaning ‘remove this part’.

If events are **not** mutually exclusive, then there’s some overlap between them - the ‘middle’ of a standard Venn diagram. Perhaps now is a good time to clarify that if two events *are* mutually exclusive, their Venn diagram-circles would not overlap at all.

First, let’s focus on the statement ‘either  $A$  or  $B$  happens’. It is pictorially equivalent to landing within either of the  $A$  or  $B$  circles, and mathematically written as the **union** of the two events,  $A \cup B$ . Let  $P(A)$  be the probability of landing within the  $A$ -circle, and  $P(B)$  the probability of landing within the  $B$ -circle. If we add them both up, say

$$P(A) + P(B),$$

then we have *accounted for the overlapping section (intersection) twice*.

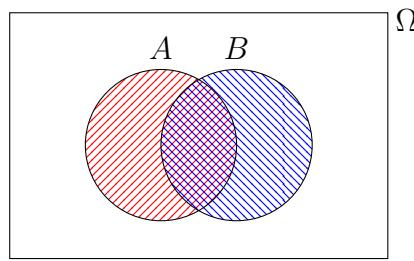


Figure 3: Accounting for both  $P(A)$  (red) and  $P(B)$  (blue) counts the *intersection* twice.

To make up for this double counting, we simply... remove one lot of it! This is precisely the ‘inclusion-exclusion’ identity,

$$P(A \cup B) = \underbrace{P(A) + P(B)}_{\text{inclusion}} - \underbrace{P(A \cap B)}_{\text{exclusion}}.$$

Inclusion-exclusion can be generalised to any number of events following the same basic principle: account for each individual event, then remove any intersections that you may have over counted.

**Exercise 1.1.** *If two events  $A$  and  $B$  are mutually exclusive, what is their intersection  $A \cap B$ ? How does the inclusion-exclusion statement change under this assumption?*

## Conditional Probability

Now we focus on a different question, namely “what is the probability of  $A$  happening given that  $B$  has happened?”

I’d like to convince you that these are far more realistic statements. It’s uncommon to know information about something without also knowing information about something else that is very closely related. Consider the following examples.

- (a) What’s the probability that it’s going to rain given that it’s very cloudy? (Certainly a lot higher than in a scenario where it’s sunny!)

- (b) What's the probability of a patient having a certain disease given that they've tested positive for it? (A lot higher than if they had tested negative, but false positives definitely exist.)

We write ‘given that’ with a vertical line,  $|$ . Thus  $A|B$  reads ‘ $A$  given that  $B$  has happened’, that is, we are *conditioning* the outcome of  $A$  on the fact that  $B$  has happened.

It is very helpful to think about this process as a *restriction of the sample space*. Remember that the sample space  $\Omega$  is the ‘set of all possible things that can happen’. If  $B$  has happened, then the ‘set of all possible things that can happen’ is within  $B$ !

Importantly, an event  $B$  may contain other individual events. For example, if I roll an even number on my dice (and rolling an even number is event  $B$ ), there are still three distinct possibilities within  $B$  in the sense that my roll could be 2, 4, or 6.

**Exercise 1.2.** *What is the probability of rolling an even number on a (fair) dice, given that the roll is strictly less than 4? How does this change from the probability of rolling an even number with no additional information on the roll?*

The idea of conditional probability on a Venn diagram is described in Figure 4. Notice that within this figure it is still possible for  $A$  to happen, it just must be contained within the intersection  $A \cap B$ .

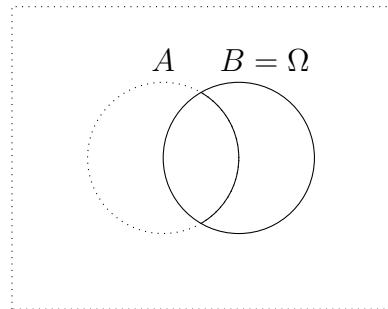


Figure 4: A Venn diagram describing an event where  $B$  has happened. The sample space is now  $B$ , so any old boundaries are now dotted to reflect this.

So what is the probability of  $A$  happening given that  $B$  has happened? It is the *fraction* of  $A$  that is contained within  $B$ , or in other words, how much of  $B$  is made up of  $A \cap B$ . Therefore

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

which can be read as ‘the probability of  $A$  happening given  $B$  has happened is the fraction of  $B$  made up of  $A$ ’. This is the statement of conditional probability.

**A note on independence.** We discussed independent events to be events that have absolutely no bearing on the other. In these cases, the probability of  $A$  happening *given that*  $B$  happens should *be the same* as  $A$  happening, completely irrespective of  $B$ . This means that, if  $A$  and  $B$  are independent, then  $P(A|B) = P(A)$ . In this scenario

$$P(A) = P(A|B) = \frac{P(A \cap B)}{P(B)},$$

so after multiplying both sides by  $P(B)$ ,

$$P(A)P(B) = P(A \cap B).$$

This is the special case of  $P(A \cap B)$  when  $A$  and  $B$  are independent.